



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genomic analysis of the multi-host pathogen *Erysipelothrix rhusiopathiae* reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution

Citation for published version:

Forde, T, Biek, R, Zadoks, R, Workentine, ML, De Buck, J, Kutz, S, Opriessnig, T, Trewby, H, van der Meer, F & Orsel, K 2016, 'Genomic analysis of the multi-host pathogen *Erysipelothrix rhusiopathiae* reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution', *BMC Genomics*, vol. 17, no. 1, 461. <https://doi.org/10.1186/s12864-016-2643-0>

Digital Object Identifier (DOI):

[10.1186/s12864-016-2643-0](https://doi.org/10.1186/s12864-016-2643-0)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

BMC Genomics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



BMC Genomics

Genomic analysis of the multi-host pathogen *Erysipelothrix rhusiopathiae* reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution

--Manuscript Draft--

Manuscript Number:	GICS-D-16-00152R3	
Full Title:	Genomic analysis of the multi-host pathogen <i>Erysipelothrix rhusiopathiae</i> reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution	
Article Type:	Research article	
Section/Category:	Prokaryote microbial genomics	
Funding Information:	Natural Sciences and Engineering Research Council of Canada (Discovery Grant)	Dr. Susan Kutz
	Natural Sciences and Engineering Research Council of Canada (Northern supplement)	Dr. Susan Kutz
	Natural Sciences and Engineering Research Council of Canada (Canada Graduate Scholarship)	Dr. Taya Forde
	University of Calgary (Eyes High)	Dr. Susan Kutz
	Killam Trusts (Izaak Walton Killam Pre-Doctoral Scholarship)	Dr. Taya Forde
	Canada North Outfitters	Dr. Susan Kutz
	Nunavut General Monitoring Program	Dr. Susan Kutz
	Nunavut Harvesters Association	Dr. Susan Kutz
	BC Oil and Gas Research and Innovation Society	Not applicable
	Weyerhaeuser Company	Not applicable
Abstract:	<p>Background: Knowledge about how bacterial populations are structured is an important prerequisite for studying their ecology and evolutionary history and facilitates inquiry into host specificity, pathogenicity, geographic dispersal and molecular epidemiology. <i>Erysipelothrix rhusiopathiae</i> is an opportunistic pathogen that is currently reemerging in both the swine and poultry industries globally. This bacterium sporadically causes mortalities in captive marine mammals, and has recently been implicated in large-scale wildlife die-offs. However, despite its economic relevance and broad geographic and host distribution, including zoonotic potential, the global diversity, recombination rates, and population structure of this bacterium remain poorly characterized. In this study, we conducted a broad-scale genomic comparison of <i>E. rhusiopathiae</i> based on a diverse collection of isolates in order to address these knowledge gaps.</p> <p>Results: Eighty-three <i>E. rhusiopathiae</i> isolates from a range of host species and geographic origins, isolated between 1958 and 2014, were sequenced and assembled using both reference-based mapping and de novo assembly. We found that a high proportion of the core genome (58%) had undergone recombination. Therefore, we used three independent methods robust to the presence of recombination to define the population structure of this species: a phylogenetic tree based on a set of conserved protein sequences, in silico chromosome painting, and network analysis. All three methods were broadly concordant and supported the existence of three distinct clades within the species <i>E. rhusiopathiae</i>. Although we found some evidence of host and geographical clustering, each clade included isolates from diverse host species and from multiple continents.</p> <p>Conclusions: Using whole genome sequence data, we confirm recent suggestions that</p>	

	E. rhusiopathiae is a weakly clonal species that has been shaped extensively by homologous recombination. Despite frequent recombination, we can reliably identify three distinct clades that do not clearly segregate by host species or geographic origin. Our results provide an essential baseline for future molecular epidemiological, ecological and evolutionary studies of E. rhusiopathiae and facilitate comparisons to other recombinogenic, multi-host bacteria.
Corresponding Author:	Taya Forde University of Glasgow UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Glasgow
Corresponding Author's Secondary Institution:	
First Author:	Taya Forde
First Author Secondary Information:	
Order of Authors:	Taya Forde
	Roman Biek
	Ruth Zadoks
	Matthew L. Workentine
	Jeroen De Buck
	Susan Kutz
	Tanja Opriessnig
	Hannah Trewby
	Frank van der Meer
	Karin Orsel
Order of Authors Secondary Information:	
Response to Reviewers:	

[Click here to view linked References](#)

Genomic analysis of the multi-host pathogen *Erysipelothrix rhusiopathiae* reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution

Taya Forde*, Taya.Forde@glasgow.ac.uk; Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada. Current address: Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, United Kingdom

Roman Biek, Roman.Biek@glasgow.ac.uk, Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, United Kingdom

Ruth Zadoks, Ruth.Zadoks@glasgow.ac.uk, Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, United Kingdom

Matthew L. Workentine, matthew.workentine@ucalgary.ca, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

Jeroen De Buck, jdebuck@ucalgary.ca, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

Susan Kutz, skutz@ucalgary.ca, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

Tanja Opriessnig, Tanja.Opriessnig@roslin.ed.ac.uk, The Roslin Institute, University of Edinburgh, Midlothian, Scotland, United Kingdom

Hannah Trewby, h.trewby.1@research.gla.ac.uk, Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, United Kingdom

Frank van der Meer, frank.vandermeer@ucalgary.ca, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

Karin Orsel, karin.orsel@ucalgary.ca, Faculty of Veterinary Medicine, University of Calgary, Calgary,
Alberta, Canada

*Corresponding author

Abstract

Background: Knowledge about how bacterial populations are structured is an important prerequisite for studying their ecology and evolutionary history and facilitates inquiry into host specificity, pathogenicity, geographic dispersal and molecular epidemiology. *Erysipelothrix rhusiopathiae* is an opportunistic pathogen that is currently reemerging in both the swine and poultry industries globally. This bacterium sporadically causes mortalities in captive marine mammals, and has recently been implicated in large-scale wildlife die-offs. However, despite its economic relevance and broad geographic and host distribution, including zoonotic potential, the global diversity, recombination rates, and population structure of this bacterium remain poorly characterized. In this study, we conducted a broad-scale genomic comparison of *E. rhusiopathiae* based on a diverse collection of isolates in order to address these knowledge gaps.

Results: Eighty-three *E. rhusiopathiae* isolates from a range of host species and geographic origins, isolated between 1958 and 2014, were sequenced and assembled using both reference-based mapping and *de novo* assembly. We found that a high proportion of the core genome (58%) had undergone recombination. Therefore, we used three independent methods robust to the presence of recombination to define the population structure of this species: a phylogenetic tree based on a set of conserved protein sequences, *in silico* chromosome painting, and network analysis. All three methods were broadly concordant and supported the existence of three distinct clades within the species *E. rhusiopathiae*. Although we found some evidence of host and geographical clustering, each clade included isolates from diverse host species and from multiple continents.

Conclusions: Using whole genome sequence data, we confirm recent suggestions that *E. rhusiopathiae* is a weakly clonal species that has been shaped extensively by homologous recombination. Despite frequent recombination, we can reliably identify three distinct clades that do not clearly segregate by

1
2
3
4 host species or geographic origin. Our results provide an essential baseline for future molecular
5
6 epidemiological, ecological and evolutionary studies of *E. rhusiopathiae* and facilitate comparisons to
7
8 other recombinogenic, multi-host bacteria.
9

10
11
12 **Key Words:** *Erysipelothrix rhusiopathiae*, genomics, Gram positive bacteria, homologous recombination,
13
14 host specificity, multi-host pathogen, population structure, whole genome sequencing
15
16

17 18 **Background**

19
20 Uncovering population structure and its determinants is essential for understanding bacterial ecology
21
22 and evolution [1]. Key questions concerning host specificity and generalism [2], global patterns of gene
23
24 flow [3], and the genetic basis for clinical disease manifestations [1] can be addressed by examining the
25
26 relationships among strains within a species. Additionally, molecular epidemiological studies rely on
27
28 having an understanding of the population structure as a framework within which to interpret the
29
30 genomic diversity of the target organism [4], and the identification of host- or geography-associated
31
32 lineages can be helpful in source attribution [5]. Whole genome sequencing is providing new
33
34 opportunities to address questions related to population structure within a phylogenetic framework;
35
36 however, this undertaking is complicated by the need to consider the potential influence of
37
38 recombination [4, 6].
39
40
41
42
43
44

45 *Erysipelothrix rhusiopathiae*, a Gram positive, facultative intracellular bacterium, is an important
46
47 opportunistic pathogen for both humans and animals. Zoonotic infections with *E. rhusiopathiae* typically
48
49 manifest as erythematous skin lesions known as erysipeloid, and tend to be occupationally associated
50
51 (e.g. slaughterhouse workers, butchers, fishermen, etc.) [7]. *E. rhusiopathiae* has been documented in a
52
53 wide range of wild and domestic species, including birds, mammals, reptiles, fish and arthropods [8].
54
55 Best known as the causative agent of swine erysipelas, *E. rhusiopathiae* can cause significant economic
56
57 losses in swine production systems due to sporadic cases of acute septicemia, subacute cutaneous
58
59
60
61
62
63
64
65

1
2
3
4 lesions, or chronic arthritis, which may be punctuated by larger outbreaks [9]; it is also among the most
5
6 common causes of carcass condemnation for swine in the United States [10]. In recent years, the
7
8 incidence of *E. rhusiopathiae* infection in swine has increased significantly in the mid-western United
9
10 States, Japan and China [10–12]. Erysipelas is also reemerging in European poultry productions, likely in
11
12 association with changes in housing systems [13, 14]. In captive marine mammals, *E. rhusiopathiae* is
13
14 known to cause serious and often life-threatening infections [15, 16], while recent die-offs involving
15
16 hundreds of muskoxen in the Canadian Arctic Archipelago have sparked interest in the potential
17
18 conservation importance of this bacterium [17]. The broad ecological and geographic distribution of *E.*
19
20 *rhusiopathiae* has been attributed to its ability to infect multiple host species which may act as healthy
21
22 carriers, in combination with its long environmental persistence [18]. However, despite its ubiquity,
23
24 importance for multiple host species including humans, and a highly variable clinical presentation, little
25
26 is known about the genetic diversity, population structure, and host specificity of *E. rhusiopathiae*.
27
28
29
30
31

32
33 Recent whole genome sequencing projects have facilitated the taxonomic classification of the genus
34
35 *Erysipelothrix* [11, 19]. A member of the phylum Firmicutes, the class Erysipelotrichia has the single
36
37 order Erysipelotrichales and family Erysipelotrichaceae, the latter comprising 10 genera [20]. Within the
38
39 genus *Erysipelothrix*, other recognized species are *E. tonsillarum*, *E. inopinata*, *E. sp. strain 1*, and *E. sp.*
40
41 strain 2 [21, 22], as well as the recently identified *E. larvae* sp. nov. [23]. *E. rhusiopathiae* and *E.*
42
43 *tonsillarum* have long been recognized as distinct species based on differences in pathogenicity,
44
45 phenotypic characteristics, and serotype groups [21, 24], supported by DNA-DNA hybridization studies
46
47 [25]. *E. inopinata* appears to have diverged prior to the split between *E. rhusiopathiae* and *E. tonsillarum*
48
49 based on 16S rRNA gene sequence phylogeny [22]. The relationship of *E. sp. strain 1* and *E. sp. strain 2*
50
51 to other *Erysipelothrix* species has not been explored.
52
53
54
55
56

57 The intraspecific classification of *E. rhusiopathiae* strains is less clearly defined, with a variety of different
58
59 tools used for categorizing isolates. Serotyping, which involves testing for agglutination with specific
60
61
62
63
64
65

1
2
3
4 antisera recognizing different peptidoglycan antigens of the cell wall [26], is one of the approaches that
5
6 has been most frequently implemented. Although it has diagnostic value, as the different *Erysipelothrix*
7
8 species have distinct sets of serotypes [27], serotyping is an inappropriate tool for inferring evolutionary
9
10 relatedness among isolates due to the frequent horizontal exchange of capsule-specific genes in many
11
12 bacterial species [28]. More recently, differences in immunogenic proteins known as surface protective
13
14 antigens (Spa) have also been used to distinguish between strains of *E. rhusiopathiae* [13, 29]. The
15
16 genetic relationship among *E. rhusiopathiae* isolates has also been examined using various comparative
17
18 genotyping methods such as pulsed-field gel electrophoresis (PFGE) [30]. However, sequence-based
19
20 typing methods, which allow for easier inter-laboratory comparison and investigation of functional
21
22 differences, have only very recently been applied to *E. rhusiopathiae*: a comparison was made among
23
24 the three *E. rhusiopathiae* whole genome sequences (two complete and one draft) available on
25
26 GenBank at the time of writing [11], and the population structure of this bacterium was examined using
27
28 multi-locus sequence typing (MLST), focusing primarily on European poultry isolates [13]. Little is known
29
30 about the importance of recombination in *E. rhusiopathiae*, although the latter study found the species
31
32 to be weakly clonal [13], suggesting recombination may contribute significantly to genetic variability in
33
34 this species. Until now there have been no large-scale genomic comparison studies to fully investigate
35
36 the genetic diversity of *E. rhusiopathiae*.
37
38
39
40
41
42
43
44

45 The objectives of this study were to describe the global genomic diversity of *E. rhusiopathiae* and to
46
47 examine its population structure while accounting for the presence of recombination, in order to
48
49 provide an essential baseline for future studies into the epidemiology and ecology of this multi-host
50
51 pathogen. A secondary objective was to examine the phylogenetic relationship among different
52
53 *Erysipelothrix* species based on whole genome sequence data.
54
55
56
57

58 Results

59
60
61
62
63
64
65

Sequencing and pan-genome statistics

Eighty-three newly sequenced *E. rhusiopathiae* isolates were included in this analysis, with representation from North America (Canada and US), South America (Argentina), Europe (Belgium, Hungary and the UK), Asia (Japan), South Africa, and Australia, with the majority originating from Canada (n=37), US (n=20) and Belgium (n=17) (see Additional file 1: Table S1 for a table of the isolates included in this study). Dates of isolation ranged from 1958 to 2014. Host species of origin encompassed swine (n=18), poultry (n=14), captive marine mammals (dolphins and beluga whales; n=8), fishes (n=5), wild birds (primarily water fowl; n=7), wild ungulates (muskoxen, moose and caribou; n=28), and a wolf. Background information regarding the clinical manifestations associated with the isolates was only available for 12 isolates (Table S1), but cases of acute and subacute septicemia, as well as *E. rhusiopathiae* isolated from skin lesions were included. Serotype was known for 16 isolates previously described in the literature (Table S1); the newly serotyped isolate VI11-2_lu was found to belong to serotype 5.

A total of 1137 core genes were present among all the *E. rhusiopathiae* isolates included in this analysis, representing 67% of the coding sequences present in the Fujisawa reference genome (Figure 1A and Additional file 2: Figure S1). A total of 512 singleton genes were identified (i.e. present in only one isolate), with an average of 6.6 new unique genes discovered per additional genome sequenced (Figure 1B). Among the core genes were the *ComEC* (membrane pore) and *dprA* (recombination mediator) genes, as well as putative Type II/IV secretion system proteins (Figure S1), the presence of which is highly suggestive of natural competency [31]. Based on the pairwise comparisons made in the program LS-BSR, no major differences in overall gene content were detected among host species or geographic origin, or among clades that were identified during analyses described below.

A high proportion of the *E. rhusiopathiae* core genome has experienced recombination

The core genome alignment generated using Parsnp was 1,049,431 base pairs, representing 58% of the reference genome. Overall there was a high degree of sequence similarity, with 99.3% pairwise identity. The BratNextGen analysis, which included all isolates, inferred that 58% of the overall core genome alignment had experienced recombination (Additional file 2: Figure S2). There were major differences among the clades in terms of the proportion of the alignment with inferred recombination (Table 1, Additional file 2: Figures S2 and S3). Both recombination detection methods estimated that a greater proportion of the genome of Clade 2 isolates has experienced recombination. BratNextGen did not detect any recombinogenic segments in Clade 1, while this was also the clade with the least recombination across the core genome as detected by Gubbins. Although a large proportion of the alignment was influenced by recombination, average recombination to mutation (r/m) rates were moderate, at 0.96, 2.18 and 0.55 in Clades 1, 2 and 3 respectively.

Phage and plasmid-related sequences were detected in several isolates

Putative prophage sequences were detected in 47 of the 86 isolates (55%), representing all host groups and geographic locations (Table S1). Up to three such sequences were detected per isolate, with a mean length of 22 KB (range: 12.5-47 KB). Twenty-eight of the isolates had phage sequences with a high level of homology ($\geq 96.5\%$ pairwise identity) with the annotated bacteriophage from the Fujisawa genome (36.5 KB; Figure S1) across at least an 8 KB segment (designated as P1 in Table S1). Two of these 28 isolates were inferred to be intact prophages by PHAST. Isolates with P1 phage sequences were present in all three clades, suggesting that the presence of these similar phage sequences is the result of either multiple introductions or ancestral acquisition with subsequent loss along various branches of the phylogeny. A second group of phage sequences (n=15; designated as P2 in Table S1) shared 95.9% pairwise identity across about 12 KB. Four of these were classified as intact by PHAST. There is a strong possibility that some of these P2 sequences have been inherited through vertical descent (e.g. in the five

Belgian swine isolates), while the other sequences were identified in unrelated isolates, indicating horizontal transfer. Eight isolates had phage sequences that clustered into a third group (P3) that had at least 12 KB of homologous sequence with 94% pairwise identity. None of these sequences were designated as intact by PHAST. These sequences all belonged to swine isolates from two distinct clusters, although not all isolates in each cluster had phage sequences detected. Among the other phage sequences detected, most shared a smaller portion of sequence homology with one of the three groups of sequences (designated as 'partial' in Table S1). Only one unique phage sequence was identified among our collection using PHAST: the 22 KB incomplete phage sequence in the poultry isolate G11. Based on BLAST comparisons, this sequence shares a 13 KB segment with an integrative conjugative element of *Streptococcus pyogenes* with 95.4% pairwise identity, while the rest of the sequence shared 96% pairwise similarity to plasmids carrying coding sequences for Type IV secretory pathway components.

The BLAST search against the publicly available *E. rhusiopathiae* plasmid sequence [GenBank: NC_002148] found that one isolate in our collection ('Ery Afrika 1', Clade 2) had a similar sequence, with 95.2% pairwise identity and 100% coverage. During the BLAST search of contigs from each isolate that did not align to the reference genome, one poultry isolate had a hit to a conjugal transfer protein, while five swine isolates had hits to plasmid sequences. Three of these, from swine from Belgium, were hits to the same plasmid sequence; searches for similar sequences among the other three Belgian swine isolates were negative.

Population structure of *E. rhusiopathiae*

The phylogenetic tree generated in PhyloPhlAn using > 400 conserved bacterial proteins, rooted to other genera of the family Erysipelotrichaceae, places *E. tonsillarum* at a position basal to *E. rhusiopathiae* and *E. sp. strain 2* in the phylogeny (Figure 2). The differentiation between clades is more easily visible when

zooming in to the within-species level (Figure 3). Consistent with this phylogeny, the analysis using ChromoPainter and fineSTRUCTURE (Figure 4 and Additional file 2: Figure S4), and network analysis implemented in SplitsTree (Figure 5) supported the existence of three distinct clades within the species *E. rhusiopathiae*. Seven isolates belonged to Clade 1 as supported by all three methods. These isolates originated from two captive marine mammals (beluga and dolphin) from the US, one dolphin from an Australian aquarium, one red wolf from the US, one fish from Japan, one caribou from Canada, and from a sheep dip from Argentina. Fourteen isolates belonged to Clade 2 as supported by all three methods. These consisted of isolates from marine mammals (n=5) from the US, Belgium and South Africa, from fish (n=3) from Japan and the US, from poultry (n=3) and swine (n=1) from Belgium, from one Canadian caribou and the ATCC19414 isolate from GenBank (host species and geographic origin unknown; see Table S1). One additional poultry isolate from Belgium (red arrow Figures 3-5) was grouped with the Clade 2 isolates by ChromoPainter, but was classified as an “intermediate” isolate between Clades 2 and 3 using phylogenetic analysis. At least five isolates clustered in this intermediate group based on all three methods: the two whole genome sequences from swine available on GenBank (SY1027 from China and Fujisawa from Japan), and two caribou and one moose isolate from Canada. By the chromosome painting method, two additional isolates clustered within this intermediate clade, both from wild birds from the US; these isolates fell within Clade 3 using the phylogenetic approach (blue and black arrows in Figures 3-5). The three isolates whose clade designation differed between the phylogenetic and chromosome painting methods all fell outside of the main clusters using network analysis (Figure 5, shown by arrows). The other 57 isolates constituted the dominant Clade 3, composed of all of the Canadian swine (n=8) and poultry (n=6) isolates tested, most of the Belgian swine isolates (6/7) and half of the Belgian poultry (4/8) isolates, three additional swine isolates (one from the US and two of unknown origin), isolates from five different North American wild bird species, isolates from North American caribou (n=11), moose (n=4) and muskoxen (n=8), and one fish isolate from Hungary.

Less concordance was observed among the different methods with respect to the relationships of isolates within the dominant Clade 3. This is illustrated in Figure 6, which shows the maximum likelihood (ML) phylogeny generated using non-recombinogenic core single nucleotide polymorphisms (SNPs) inferred by Gubbins, with clusters supported by either PhyloPhlAn or fineSTRUCTURE superimposed. In general there was stronger support for more recent nodes based on higher bootstrap values in the ML tree, as well as more frequent support by the other two methods. No temporal signal was detected in the root-to-tip analysis of Clade 3 isolates using Path-O-Gen (correlation < 0.001).

A consistent relationship was found between Spa-type and clade: all isolates in Clade 1 had a single copy of the *SpaB* gene, while all other *E. rhusiopathiae* isolates had a single copy of *SpaA* (Figure 3). Conversely, among those isolates whose serotype had been defined, significant homoplasy was observed (Additional file 2: Figure S5). Serotype 5 isolates were found in all three clades, serotype 2 isolates were found in Clades 2 and 3, and serotype 1a isolates fell into both intermediate and Clade 3.

Limited evidence for host or geographic specificity

All three clades included isolates from various host species and from multiple continents. However, all marine mammal isolates (n=8) fell into Clades 1 or 2 and none within the dominant Clade 3, while no swine or poultry isolates were found to belong to Clade 1. Within Clade 3, clustering by host species, geographic location, and submitting laboratory were all found to be significant using the program BaTS (Additional file 3: Table S2). Although not fully supported by the other two methods, the phylogenetic tree based on non-recombinant SNPs (Figure 6) grouped the majority of livestock isolates within Clade 3 (21/27 swine and poultry isolates) into a single sub-group with 96% bootstrap support, whereas the majority of isolates of wildlife origin fell into a separate sub-group within this clade.

Discussion

1
2
3
4 **Homologous recombination has occurred extensively throughout the core genome of *E.***
5
6 ***rhusiopathiae***
7

8
9 This is the first large-scale genomic study of the multi-host pathogen *E. rhusiopathiae*. Our results
10 indicate that homologous recombination plays an important role in generating diversity within this
11 species, confirming recent findings by an MLST study [13]. Based on our data, several mechanisms may
12 contribute to recombination in this species: the uptake of genetic material from the external
13 environment, phages, and plasmids. Bacteria capable of importing foreign DNA across the cell envelope
14 (transformation) are said to be ‘naturally competent’ and possess specific machinery for this task [31].
15
16 The presence of the *ComEC* (membrane pore) and *dprA* (recombination mediator) genes within the core
17 genome, as well as coding sequences for putative Type II/IV secretion system proteins, is highly
18 suggestive that *E. rhusiopathiae* has this capacity [31, 32]. Transduction and conjugation may also be
19 important mechanisms for horizontal gene transfer in *E. rhusiopathiae*, since phage-related sequences
20 were abundant among the isolates in our study and some plasmid-related sequences were also
21 identified, despite the fact that the sensitivity for detecting such sequences from draft assemblies may
22 be low [33]. In previous studies, up to a third of isolates tested were found to harbor plasmids [34, 35],
23 and several mobile genetic elements in the available reference genomes bore signatures of having
24 plasmid or prophage origin [11]. To our knowledge, this is the first study to have examined the
25 prevalence and diversity of phage-related sequences in a large collection of *E. rhusiopathiae* isolates.
26
27 Although limited data were available regarding the clinical presentation associated with many of the
28 isolates, no clear association was found between the presence of phage sequences and pathogenicity.
29
30 The presence of particular phage sequences did not appear to be a phylogenetically informative marker
31 for epidemiological inference, since related phage sequences were found throughout the different
32 clades, and in some cases, phage sequences were not conserved within groups of highly similar isolates.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 Similarly, the inconsistent presence of plasmid sequences among closely related isolates suggests that
5
6 conjugation may occur frequently.
7
8

9
10 Although a high proportion of the core genome has experienced recombination events, the *r/m* rates we
11
12 estimated for each Clade (between 0.55 and 2.18) were only moderate in comparison to other bacterial
13
14 species [36]. This, along with the high pairwise identity (99.3%) found within the core genome, suggests
15
16 that within the *E. rhusiopathiae* core genome, alleles are highly conserved, although its organization is
17
18 not. This may explain the lack of concordance found between MLST and PFGE [13], since this latter
19
20 technique is sensitive to structural rearrangements. Within each clade, a larger proportion of the core
21
22 genome was estimated to have experienced recombination based on the Gubbins output in comparison
23
24 to BratNextGen. Gubbins is based on the method initially cited in Croucher et al. [37], and has since
25
26 been employed in several other studies of bacterial pathogens [38, 39]. Inference of recombinant
27
28 segments is based on the detection of areas of higher SNP density in comparison to the background
29
30 threshold level. Since areas of elevated mutations are not necessarily specific for recombination and
31
32 could instead represent regions of the genome with naturally higher mutation rates or that are under
33
34 less purifying selection, this could be a potential source of false positives using this method [40].
35
36 BratNextGen, on the other hand, uses a Bayesian change-point clustering model to detect evolutionarily
37
38 distinct lineages; these clusters are taken as the putative ancestral origins when estimating
39
40 recombination probabilities in each isolate. This program has been frequently cited in intraspecific
41
42 bacterial population studies based on whole genome data [41, 42]. Despite the differences in the extent
43
44 of recombination found between the two programs used, both suggest that recombination has occurred
45
46 most frequently in Clade 2, where the *r/m* rate was two to four times higher compared to the other
47
48 clades and a greater proportion of the genome was found to have been implicated. Clade-associated
49
50 differences in recombination rates have been previously observed in other bacterial species such as
51
52 *Listeria monocytogenes* [43].
53
54
55
56
57
58
59
60
61
62
63
64
65

The population structure of *E. rhusiopathiae* was consistent across multiple inference approaches

Given the uncertainties associated with whole genome phylogenetic reconstruction in the presence of recombination, we took three conservative, independent approaches for inferring the population structure of *E. rhusiopathiae*. The use of a set of conserved protein sequences has been shown to be robust to horizontal gene transfer [44]. Chromosome painting does not rely on phylogenetic inference but rather reconstructs the chromosome haplotype of “recipient” individuals as a composition of recombination-derived segments from the other “donor” individuals [45]. Finally, network analysis provides a means by which to visualize alternative phylogenetic histories in organisms that are not strictly clonal [46]. Using these three approaches, good concordance was found for assigning isolates into clades. The fact that these methods were based on input from both *de novo* genome assembly (conserved protein tree and network analysis) and a reference-based mapping approach (chromosome painting), provides support for the robustness of these results.

Clade 1 could be distinguished from the other two clades and intermediate isolates by the presence of a *SpaB* gene, as opposed to *SpaA* found in the other isolates. These surface protective antigens are potent immunogens with potential relevance for vaccine development [29]. *SpaA* is among the more well-characterized genes that have been proposed to be associated with *E. rhusiopathiae* pathogenicity [19], with a role in endothelial adherence [47] and resistance to phagocytosis [48]. The *SpaB* gene shares approximately 60% amino acid similarity with *SpaA* and is antigenically distinct [49]. It was previously found that *E. rhusiopathiae* strains of aquatic origin were able to express more than one Spa-type [29], although an underlying genetic basis for this was not observed among the isolates in our collection. Not surprisingly, serotype was not associated with the phylogenetic relatedness among isolates [50, 51]. Since homoplasmy of serotype is likely commonplace (Figure S5), it should not be used to infer evolutionary relatedness among isolates, and its use as an epidemiological tool (e.g. for deciding

whether multiple isolates belong to the same outbreak) would require additional information on the expected frequency of serotype switching.

Despite removing recombinogenic sites prior to constructing the phylogeny of the dominant Clade 3, basal relationships among these isolates were difficult to elucidate, with low bootstrap values obtained for most of the deeper nodes (Figure 6). It is possible that not all recombination events were detected and that this still had a confounding influence; a PHI-test revealed a strong signal of recombination even in this curated SNP set. The lack of a detectable clock-like signal could be partially due to residual recombinogenic segments in this dataset [4], although given the lack of correlation between the date of isolation and root-to-tip divergence, it seems probable that additional factors are contributing.

Hypothetically, highly variable generation times (e.g. little bacterial replication during environmental phases or carriage in comparison to during active infection) could be a plausible explanation for the lack of temporal structure. Failing to calibrate the molecular clock, we were unable to estimate a global or clade-specific substitution rate for *E. rhusiopathiae* based on whole genome sequence data.

***E. rhusiopathiae* shows limited host association**

Bacteria show remarkable variability in the extent to which they specialize to colonize specific host species, the determinants of which are thought to be the result of complex host-pathogen interactions [52]. Many bacterial species have fairly distinct host-adapted strains [2, 53], while other bacterial species are characterized by a mix of host-adapted and generalist sub-types [42, 54]. The determinants of host specificity can be seen along a broad continuum from changes in a single amino acid residue, to changes in gene content (presence/absence) or the acquisition of genomic islands [55]. In our study, no consistent differences in gene content were found between isolates from different host species, geographic locations or clades.

1
2
3
4 It is known from experimental infection studies that multiple host species are susceptible to the same
5
6 strains of *E. rhusiopathiae*; isolates from one host species that are subsequently inoculated into another
7
8 species often result in infection and clinical disease in the recipient species [56], although susceptibility
9
10 varies [57]. We found that all three clades comprised isolates from a variety of taxonomically very
11
12 different host species. It is, therefore, probable that all *E. rhusiopathiae* lineages are infectious for a
13
14 wide range of species, although possibly with varying levels of infectiousness and pathogenicity. Our
15
16 data also contained strong indications of cross-species transmission given that some isolates from
17
18 wildlife were nested within lineages dominated by domestic hosts and vice versa (Figures 3 & 6),
19
20 suggesting host or ecological predilection exists, but not strict host specificity. Similar findings were
21
22 reported using MLST, wherein the dominant clonal complex encompassed isolates from multiple host
23
24 species, including poultry, pigs, sheep and humans [13].
25
26
27
28
29
30

31 Although the different clades do not appear to be strictly limited to specific hosts, the observation that
32
33 isolates from captive marine mammals fell exclusively into Clades 1 and 2 is significant ($p < 0.0001$),
34
35 especially given that these are from aquariums distributed globally. Whether this finding is associated
36
37 with differences in susceptibility of marine mammals to these strains, or that these strains are
38
39 differentially distributed due to adaptations to marine environments warrants further investigation. Both
40
41 Clade 1 and Clade 2 *E. rhusiopathiae* isolates were associated with fatal infections in marine mammals
42
43 (Table S1). However, no Clade 1 isolates, and therefore no strains carrying the *SpaB* gene, were found in
44
45 any of the swine or poultry isolates in our collection. Since it has been previously shown that strains of
46
47 marine origin carrying the *SpaB* gene can experimentally cause at least mild to moderate lesions in
48
49 swine [56], the lack of Clade 1 strains in swine and poultry is more likely due to lack of exposure than a
50
51 lack of pathogenic potential.
52
53
54
55
56
57

58 Within the dominant Clade 3, although clustering by host species was detected, equally strong support
59
60 for clustering based on geographic origin and source laboratory was found (Table S2), suggesting that
61
62
63
64
65

1
2
3
4 these associations cannot be separated from sampling bias related to the opportunistic nature of our
5
6 isolate collection. This represents a common challenge in phylogenetic studies of pathogens [58].
7
8 Further examination into host and geographic clustering within this clade using a broader set of sample
9
10 sources, including more detailed metadata in order to rule out clustering due to epidemiologically linked
11
12 isolates, would be of value. Despite this limitation, there appears to be a non-random segregation of
13
14 Clade 3 isolates between those of livestock and wildlife origin (Figure 6); swine and poultry isolates from
15
16 three independent sources of both North American and European origin tended to cluster
17
18 independently of those from northern ungulates. Among the livestock isolates, there was evidence of
19
20 regular cross-continental exchange.
21
22
23

24
25 Although efforts were made to include a global collection of isolates from a representative variety of
26
27 host species, most isolates originated from North America and Europe, and some important hosts
28
29 including humans, reptiles and arthropods were not part of our collection. The characterization of
30
31 additional isolates, particularly from under-represented continents and host species, will help build our
32
33 understanding of the global population structure of *E. rhusiopathiae*.
34
35
36

37 38 **Whole genome sequencing supports the current division among *Erysipelothrix* species**

39
40 We have shown *E. sp.* strain 2 to be phylogenetically distinct from both *E. rhusiopathiae* and *E.*
41
42 *tonsillarum* based on the conserved protein phylogeny. Its distinction from *E. rhusiopathiae* is further
43
44 supported by the fact that its level of sequence similarity was insufficient for mapping of reads to the
45
46 full genome sequences of either Fujisawa or SY1027, or for inclusion in the core genome alignment
47
48 generated through Parsnp, which requires $\geq 97\%$ average nucleotide identity. Based on the conserved
49
50 protein phylogeny, *E. rhusiopathiae* and *E. sp.* strain 2 appear to have evolved from the less pathogenic
51
52 species *E. tonsillarum*. This relationship was not observed in previous phylogenetic studies that
53
54 combined the output from multiple phylogenetic approaches [20]. Further investigation into the genetic
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 differences separating *E. rhusiopathiae* and *E. sp. strain 2* from *E. tonsillarum* could potentially provide
5
6 insights into how these species have acquired a higher level of pathogenicity. The recent detection of
7
8 *Erysipelothrix* spp. in the subsurface biosphere may represent an interesting opportunity for examining
9
10 the evolution of this genus in a broader context [59].
11
12
13

14 15 **Conclusions**

16
17 Using different approaches, we were able to confidently determine the population structure of the
18
19 multi-host pathogen *E. rhusiopathiae*, despite the important role that recombination has had in its
20
21 evolutionary history. Evidence was found to suggest that novel DNA may be acquired by this bacterium
22
23 through transformation, transduction and conjugation, lending plasticity to the genome. The species
24
25 comprises three major clades that are found across multiple continents and host species representing
26
27 both livestock and wildlife, with some indication that clades or subclades may differ in their host
28
29 predilection and recombination rate. Epidemiological inference was hampered by the opportunistic
30
31 nature of the isolate collection available for genome sequencing and future studies would benefit from
32
33 targeted sample collection. Nonetheless, our results provide an essential framework for supporting
34
35 future in-depth epidemiological and evolutionary studies involving this species and comparative studies
36
37 with other recombinogenic, multi-host bacteria.
38
39
40
41
42
43

44 45 **Methods**

46 47 **Bacterial isolates**

48
49 In order to examine the intraspecific genomic diversity of *E. rhusiopathiae*, isolates were
50
51 opportunistically collected so as to encompass a broad range of geographic locations, host species,
52
53 clinical manifestations, and years of initial isolation. *E. rhusiopathiae* isolates, as well as isolates from
54
55 other *Erysipelothrix* spp. (*E. tonsillarum* and *E. sp. strain 2*) were kindly provided by various collaborators
56
57 (Table S1). Upon receipt, isolates were sub-cultured onto Columbia Agar (CA) with 5% sheep blood (BD-
58
59
60
61
62
63
64
65

Canada, Mississauga, ON, Canada) for morphological characterization; if colony morphology was characteristic of *E. rhusiopathiae* (clear to pale blue in color, circular, small diameter, often with alpha hemolysis [18]), a single colony was re-streaked to obtain a clonal population for DNA extraction. In addition to these archived isolates, further isolates were obtained through sample testing associated with wildlife health surveillance projects. Various tissue samples from wild ungulates (moose, caribou and muskoxen) were selectively cultured for *E. rhusiopathiae*: 2 g of tissue were mechanically homogenized in 20 ml of brain heart infusion (BHI) broth with 5% serum using a Stomacher 80 Biomaster (Seward, Port Saint Lucie, FL, USA), incubated overnight at 37°C with 5% CO₂, followed by 48 hour incubation in selective medium containing kanamycin (40 µg/ml), neomycin (50 µg/ml) and vancomycin (25 µg/ml) [60, 61] and sub-culture to agar plates of the same selective medium for 48-72 hours at 37°C with 5% CO₂. Colonies were further sub-cultured on CA plates as described for the other isolates.

DNA extraction and PCR confirmation

DNA was extracted from clonal populations by suspending 1 loopful of colonies in 200 µl of phosphate buffered saline (PBS), and then using the DNeasy Blood and Tissue Kit (Qiagen, Mississauga, ON, Canada) following manufacturer's instructions, with the exception that DNA was eluted in a smaller volume (50 µl) of the provided elution buffer to yield a higher DNA concentration. Extracted DNA was confirmed to be from *E. rhusiopathiae* by qPCR using previously described primers and a species-specific probe targeting the 3' non-coding region of the rRNA gene cluster [27]. The 20 µl PCR reaction consisted of 10 µl of TaqMan Fast Advanced Master Mix (Applied Biosystems, Carlsbad, CA), 10 pmol of each primer, 1 pmol of probe, and 2 µl of template DNA. Reactions were performed using a CFX96 thermocycler (Bio-Rad, Mississauga, Ontario, Canada) with the following cycling conditions: 50°C for 2 min, 95°C for 20 sec, then 40 cycles of 95°C for 3 sec and 57°C for 30 sec. *E. tonsillarum* isolates were also confirmed by probe-based qPCR using a different probe with this same set of primers and reaction conditions [27].

The identity of *E. sp.* strain 2 was confirmed *in silico* by identifying the specific primer sequences for the *E. sp.* strain 2 23S rRNA gene in the draft assembly [62].

Library preparation and sequencing

Library preparation was performed using the Nextera XT v2 kit (Illumina, San Diego, CA) following manufacturer's instructions, including a PhiX control spiked in at 1%. Multiplex sequencing was performed on an Illumina MiSeq machine at the University of Calgary, resulting in 250 base pair paired-end reads.

Read mapping and variant detection

The program ConDeTri was used to trim raw reads, extract those of high quality, and remove duplicates [63] (see Additional file 2: Figure S6 for a schematic diagram of the analysis pipeline, and Additional file 4: File S1 for specific commands and parameters used for each program). Unique trimmed reads from each isolate were mapped against the complete *E. rhusiopathiae* Fujisawa genome [Genbank:NC_015601] using BWA-MEM with default settings [64]. Sequencing and mapping parameters (number of reads, number of mapped reads, mean coverage, mean mapping quality) were extracted from .bam files using Qualimap [65] (Table S1). All sequenced isolates had a minimum mean coverage of 9X, with an average mean coverage of 33X. Median mapping quality was consistently high, with a mean value of 57.

Variants were detected using the mpileup command in SAMtools [66]. A list of high quality SNP sites across all *E. rhusiopathiae* genomes was generated using custom python scripts (written by HT; available upon request) that filtered on base quality, mapping quality, read depth, and heterozygosity. A site was included in the variant list if it had consensus base quality ≥ 40 , mapping quality ≥ 40 , at least 3 reads mapping to that site on each strand, and the majority base present in $>95\%$ of reads at that site for at least one isolate. Sites were excluded where more than one alternate allele was found since these

positions were believed to be less phylogenetically reliable due to possible sequence saturation or mapping error. Alleles at each SNP site were called across all isolates if the consensus base quality was ≥ 30 , mapping quality was ≥ 30 , a minimum of 2 reads were present on each strand, and $>95\%$ of reads supported the same allele in the isolate. Additional whole genome sequences of *E. rhusiopathiae* available on GenBank at the time of writing, strain SY1027 [Genbank:NC_021354] and the draft sequence of the strain ATCC19414 [Genbank:NZ_ACLK000000000] were included by simulating reads using the wgsim script in SAMtools and mapping these back to the reference genome in the same way as the other isolates for allele calling. Mobile and repetitive elements were identified in the annotated reference genome, as well as using the repeat-match command in MUMmer [67] and were removed from the variant list. A total of 32,148 high quality unique variant sites compared to the Fujisawa reference genome were identified across all 85 *E. rhusiopathiae* isolates (including the two GenBank sequences) once SNPs within repetitive regions and mobile elements were removed. A list of sites where alleles were correctly called in 100% of isolates ($n=6078$) was then generated. A total of 142 *E. rhusiopathiae* isolates were sequenced [Genbank BioProject: PRJNA288715], however in cases where multiple isolates were considered to be epidemiologically linked based on metadata and sequence similarity (i.e. <15 high quality SNP differences across the whole genome), only the isolate with the highest coverage was selected for inclusion in this analysis.

***De novo* assembly**

Unique, trimmed reads were assembled using SPAdes [68] applying a k-value of 55; this assembler was selected as the best compromise between contig number/L50 and fewer misassemblies based on metrics generated for candidate assemblers using QUAST [69]. SPAdes was run using the built-in BayesHammer for further read error correction [70] and Mismatch Corrector, a post processing tool that uses BWA. The PAGIT suite of programs was used to order and extend contigs and correct errors in the consensus sequence [71]. Output assembly metrics pre-PAGIT (number of contigs, N50) and post-PAGIT

(number of scaffolds and GC %) were determined using QUAST (Table S1). The average number of scaffolds per isolate after the PAGIT improvement pipeline was 13 (range: 4-89). The average GC% ranged from 35.9- 36.5%, with the mean across all isolates being 36.3%.

Annotation

RAST (Rapid Annotation using Subsystem Technology) was used to annotate and predict coding sequences in each isolate [72]. Putative prophage sequences were detected using PHAST [73]. To identify potential plasmid-associated sequences, contigs that were not aligned to the reference genome in PAGIT were searched against the NCBI database using Megablast. Additionally, a BLAST search of the *E. rhusiopathiae* plasmid sequence available on GenBank was performed against all assemblies. Hits to known plasmids or to genes encoding conjugal transfer proteins were considered suggestive of the presence of plasmids. To determine the Spa-type of each isolate, a custom BLAST database including representative sequences from the three recognized Spa types (A, B and C) [49] was searched against each *E. rhusiopathiae* isolate using Geneious version 7.1.8 [74]. The serotype of isolate VI11-2_lu was determined using methods previously described [56]. The genetic basis for differences in serotype has not been previously described in the literature, therefore it was not possible to determine the serotype of unknown isolates based on the genomic sequence data.

Comparative genomics

Pan genome statistics were generated using LS-BSR [75] and visualized using PanGP [76]. To improve sensitivity in detecting core proteins that might be missed due to inferior assembly quality in isolates with a lower depth of coverage, isolates with coverage less than 15X (n=8) were excluded from this analysis. The LS-BSR script compare_BSR.py was used to look for differences in gene content between groups of isolates. Specifically, pairwise comparisons in gene content (genes present in all members of one group that are absent in all members of a second group) were made among the different Clades (as

determined by PhyloPhlAn and fineSTRUCTURE), between isolates originating from specific host species (swine, poultry, marine mammal and muskox isolates compared to all other isolates), as well as geographic origin (North American vs. European). Unfortunately, given the lack of metadata on clinical manifestation associated with our collection of isolates, it was not possible to assess genetic differences that might be associated with pathogenicity in this study. The core genetic content was visualized in GView using isolates with coverage $\geq 15X$ [77].

Tests for recombination

The program Parsnp [78] was used to generate a core genome alignment (i.e. conserved orthologous regions present in all included genomes). Input for this alignment was the 83 PAGIT-improved *E. rhusiopathiae* *de novo* assemblies, as well as the three *E. rhusiopathiae* sequences available from GenBank. Putative prophage sequences detected with PHAST were masked using the BEDTools maskfasta command [79] prior to creating the alignment. This core nucleotide alignment was used as the input for both BratNextGen [80] and Gubbins [40]. Since Gubbins is best suited for detecting recombination in closely-related groups of isolates [40], separate analyses were run for each clade of isolates as later determined using PhyloPhlAn and fineSTRUCTURE. BratNextGen was run setting the hyperparameter α to 4 and using 20 iterations of the recombination estimation algorithm. The statistical significance was estimated using 100 permutations of the algorithm, setting significance at $\alpha = 0.05$. Gubbins was run within the publicly available virtual machine using default settings. Recombination to mutation rates (r/m; number of SNPs in recombinogenic segments: number of SNPs inferred to be the result of spontaneous mutation) were calculated from Gubbins output by taking the sum of mutations inside and outside recombinant segments along all internal branches leading to each terminal node of the output tree, starting from the node of the inferred common ancestor of that clade.

Examination of the population structure

Since we found that recombination has played an important role in generating the diversity observed within the species *E. rhusiopathiae*, three techniques that are robust to recombination were used to examine the population structure. First, PhyloPhlAn was used to estimate a phylogenetic tree based on a set of > 400 conserved protein sequences common to most bacteria [44]. The amino acid fasta files generated using RAST were used as the input for the program. The following isolates were included: 83 *de novo* and three *E. rhusiopathiae* strains available on GenBank, three *de novo* assembled *E. spp.* isolates (Table S1), as well as the draft sequence of *E. tonsillarum* [NZ_AREO000000000]. As an outgroup, draft whole genome sequences of a selection of members from closely related genera [20] were retrieved from GenBank. These were of *Holdemania filiformis* DSM 12042 [NZ_ACCF000000000], *[Clostridium] spiroforme* DSM 1552 [NZ_ABIK000000000], *Erysipelatoclostridium ramosum* DSM 1402 [NZ_ABFX000000000], *Solobacterium moorei* F0204 [NZ_AECQ000000000], and *Bulleidia extructa* W1219 [NZ_ADFR000000000]. A circularized version of this tree was generated in the Interactive Tree of Life (iTOL) v3.0 to illustrate the relationship among the *E. rhusiopathiae* isolates and their host species and geographic origin [81].

ChromoPainter and fineSTRUCTURE were implemented as an alternative method for examining population structure without relying on phylogenetic inference [45]. This analysis assesses each genome (recipient) as the sum of segments that could have been received from any of the other genomes in the analysis (donors). These inferred shared segments are summarized in a 'co-ancestry matrix' through Principal Component Analysis which is then used as the input for Bayesian Markov chain Monte Carlo (MCMC) model-based clustering. We used the SNP alignment obtained from the reference-based mapping approach as the input for ChromoPainter, including all 86 *E. rhusiopathiae* isolates. Instructions in the example provided on the official program web page were followed [82]. A uniform recombination map was generated using a perl script available with this program, and was used in ChromoPainter's E-

M procedure for estimating the parameters of effective population size and global mutation rate (estimated to be 2379.89 and 1.17 E-3 respectively). These parameters were then used to run ChromoPainter, followed by ChromoCombine which calculates the variance expected in the data, needed for running fineSTRUCTURE. This was run using 1,000,000 MCMC iterations, half of which were burn-in, sampling every 250 iterations. A heat map representing the inferred relationships was generated using the fineSTRUCTURE graphical user interface.

Finally, a phylogenetic network was inferred using Neighbor-Net [46], implemented within SplitsTree using a core nucleotide alignment of all 86 *E. rhusiopathiae* genomes without masked phage sequences generated in Parsnp. To achieve greater resolution among the isolates of Clade 3 which was found to be the dominant clade, the SNP sites determined to be outside of recombinant segments by Gubbins (n=7580) were used to generate a maximum likelihood tree of the isolates within this clade using PhyML [83], using a generalized time-reversible (GTR) model of nucleotide substitution with a gamma distribution, and performing 1000 bootstrap replicates. To test whether a temporal signal could be detected in Clade 3 isolates once recombinant segments had been removed, this maximum likelihood tree was examined in Path-O-Gen, which looks for correlation between year of isolation and root-to-tip divergence [84]. Clustering within Clade 3 by host species (swine/poultry/wild birds/wild ungulates/other), geographic origin (Europe, central North America, Northern Canadian provinces, Arctic, other/unknown) or submitting laboratory was examined in BaTS [85]. BaTS analysis was run on a set of trees generated in MrBayes [86], using the curated set of non-recombinant SNPs from Gubbins as the input alignment, performing 10 million iterations, sampling every 1000 samples, and taking the first half as burn-in. A Fisher's exact test was used to assess the statistical significance of the lack of Clade 3 isolates found among marine mammals.

Declarations

Ethics approval

Sample collections from wild ungulate carcasses were approved by the University of Calgary Animal Care Committee (AC13-0072) which adheres to the guidelines of the Canadian Council on Animal Care.

Availability of data

All raw sequence reads (Table S1) are available on the NCBI Sequence Read Archive [BioProject PRJNA288715].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant and Northern supplement (Kutz) and Canada Graduate Scholarship (Forde); University of Calgary Eyes High; Izaak Walton Killam Pre-Doctoral Scholarship; Canada North Outfitters, Nunavut General Monitoring Program; Nunavut Harvesters Association; BC Oil and Gas Research and Innovation Society; and Weyerhaeuser Co. Ltd.

Authors' contributions

TF performed the laboratory work, sequencing and analyses. TF, RB, RZ and KO drafted the manuscript. TF, RB, RZ, JDB, SK, HT, FvdM and KO contributed to the experimental design. RB, RZ, MLW, HT and JDB contributed analytic ideas. MLW assisted with bioinformatics. TO performed the serotyping. HT wrote the python scripts. All authors provided intellectual input and read and approved the final manuscript.

Acknowledgments

We would like to thank all those who kindly contributed isolates and associated metadata to this study: Filip Boyen, Freddy Haesebrouck, Geraldine Lacave, Sarah Wambacq and Eric Cox from Ghent University,

1
2
3
4 Serge Messier from Université de Montréal, David Blehert and Kim Miller from the US Geological Survey,
5
6 Bill van Bonn and Brigita Harris from the John G. Shedd Aquarium, Jackie Gallant of Gallant Laboratories,
7
8 and the Canadian Wildlife Health Cooperative (CWHC). We would also like to thank the many people
9
10 involved in sample collections from caribou, moose and muskoxen: Layne Adams from the US Geological
11
12 Survey, Sylvia Checkley and Bryan Macbeth of University of Calgary, Mathieu Dumond, Amanda
13
14 Niptanatiak, Shane Sather, Jorgen Bolt, and Rob Harmer of Government of Nunavut, Marsha Branigan,
15
16 Tracy Davison and Brett Elkin of Government of Northwest Territories (NWT), local hunters in both
17
18 Nunavut and NWT, Cait Nelson and Helen Schwantje of Government of British Columbia, Laura Finnegan
19
20 and Gord Stenhouse of the Foothills Research Institute, and members of the CWHC. Finally, we would
21
22 like to thank Christina Ahlstrom, Adam Chernick, Sohail Naushad, James Wasmuth, Richard Orton,
23
24 Kathryn Holt and Kelly Wyres for their advice and technical support. This research was enabled in part
25
26 by computing and technical support provided by WestGrid (www.westgrid.ca) and Compute Canada
27
28 (www.computecanada.ca).
29
30
31
32
33
34

35 36 **Additional files**

37
38 **Additional file 1: Table S1.** Excel spreadsheet of the *Erysipelothrix* spp. isolates used in this study,
39
40 including metadata and assembly statistics.
41
42

43 44 **Additional file 2: Supplementary Figures**

45
46 **Figure S1.** *Erysipelothrix rhusiopathiae* core genome. Core genes were plotted against the *E.*
47
48 *rhusiopathiae* Fujisawa reference genome using GView, filtering out low-complexity sequences (e.g.
49
50 repetitive regions). Bacteriophage sequences in the annotated reference genome and core genes
51
52 associated with bacterial competence are highlighted. Publicly available *E. rhusiopathiae* isolates and *de*
53
54 *novo* assembled isolates whose average depth of coverage was greater than 15X were included in this
55
56 analysis.
57
58
59
60
61
62
63
64
65

Figure S2. Recombinant fragments estimated in the *Erysipelothrix rhusiopathiae* core genome using BratNextGen. Presence of the same color block across multiple isolates within a column represents acquisition of the same recombinant segment; otherwise colors are arbitrary.

Figure S3. Recombinant fragments in Clades 1, 2 and 3 respectively, estimated in Gubbins. Red blocks are recombinant fragments that have been inherited by multiple isolates, while blue fragments are unique to that isolate.

Figure S4. Population subgroup assignment of *Erysipelothrix rhusiopathiae* isolates during chromosome painting using ChromoPainter/fineSTRUCTURE as shown in Figure 4.

Figure S5. Homoplasmy associated with *Erysipelothrix rhusiopathiae* serotyping. This figure shows examples of homoplasmy in serotypes 1a, 2 and 5, based on the same phylogenetic tree shown in Figure 3. Labelled arrows show the locations of isolates of the different serotypes. Serotype was previously determined for all isolates shown here except VI11-2_Lu, which was serotyped in this study.

Figure S6. Data analysis pipeline.

Additional file 3: Table S2. Excel spreadsheet presenting cluster statistics for host species, geographic location, and submitting laboratory, as output from the program BaTS.

Additional File 4: File S1. Program versions and specific command line options.

References

1. Wilson DJ. Insights from genomics into bacterial pathogen populations. PLoS Pathog. 2012;8:e1002874.

2. Pannekoek Y, Dickx V, Beeckman DSA, Jolley KA, Keijzers WC, Vretou E, et al. Multi locus sequence typing of *Chlamydia* reveals an association between *Chlamydia psittaci* genotypes and host species. PLoS ONE. 2010;5:e14179.
3. Chin C-S, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, et al. The origin of the Haitian cholera outbreak strain. New Engl J Med. 2010;364:33–42.
4. Croucher NJ, Harris SR, Grad YH, Hanage WP. Bacterial genomes in epidemiology--present and future. Philos Trans R Soc Lond, B, Biol Sci. 2013;368:20120202.
5. Muellner P, Pleydell E, Pirie R, Baker M, Campbell D, Carter P, et al. Molecular-based surveillance of campylobacteriosis in New Zealand – from source attribution to genomic epidemiology. Euro Surveill. 2013;18:pii:20365.
6. Martin DP, Beiko RG. Genetic recombination and bacterial population structure. In: Robinson DA, Falush D, Feil EJ, editors. Bacterial population genetics in infectious disease. Hoboken: Wiley-Blackwell; 2010. p. 61–85.
7. Veraldi S, Girgenti V, Dassoni F, Gianotti R. Erysipeloid: a review. Clin Exp Dermatol. 2009;34:859–862.
8. Leighton FA. Erysipelothrix infection. In: Williams ES, Barker I, editors. Infectious diseases of wild mammals. 3rd ed. Hoboken: Wiley-Blackwell; 2008. p. 491–493.
9. Opriessnig T, Wood RL. Erysipelas. In: Zimmerman JJ, Karriker LA, Ramirez A, Schwartz KJ, Stevenson GW, editors. Diseases of swine. 10th ed. Hoboken: John Wiley & Sons; 2012. p. 750–759.
10. Bender JS, Irwin CK, Shen H-G, Schwartz KJ, Opriessnig T. *Erysipelothrix* spp. genotypes, serotypes, and surface protective antigen types associated with abattoir condemnations. J Vet Diagn Invest. 2011;23:139–142.

- 1
2
3
4 11. Kwok AH, Li Y, Jiang J, Jiang P, Leung FC. Complete genome assembly and characterization of an
5 outbreak strain of the causative agent of swine erysipelas – *Erysipelothrix rhusiopathiae* SY1027. BMC
6 Microbiol. 2014;14:176.
7
8
9
10
11
12 12. To H, Sato H, Tazumi A, Tsutsumi N, Nagai S, Iwata A, et al. Characterization of *Erysipelothrix*
13 *rhusiopathiae* strains isolated from recent swine erysipelas outbreaks in Japan. J Vet Med Sci.
14 2012;74:949–953.
15
16
17
18
19
20 13. Janßen T, Voss M, Kühl M, Semmler T, Philipp H-C, Ewers C. A combinational approach of multilocus
21 sequence typing and other molecular typing methods in unravelling the epidemiology of *Erysipelothrix*
22 *rhusiopathiae* strains from poultry and mammals. Vet Res. 2015;46:84.
23
24
25
26
27
28 14. Eriksson H, Nyman A-K, Fellström C, Wallgren P. Erysipelas in laying hens is associated with housing
29 system. Vet Rec. 2013;173:18.
30
31
32
33 15. Boehm J, Lacave G, Patterson R, editors. Proceedings of the first international workshop on
34 erysipelas in cetaceans. Chicago, Illinois, USA; 2000. 101 pp.
35
36
37
38
39 16. Lacave G, Cox E, Hermans J, Devriese L, Goddeeris BM. Induction of cross-protection in mice against
40 dolphin *Erysipelothrix rhusiopathiae* isolates with a swine commercial vaccine. Vet Microbiol.
41 2001;80:247–253.
42
43
44
45
46 17. Kutz S, Bollinger T, Branigan M, Checkley S, Davison T, Dumond M, et al. *Erysipelothrix rhusiopathiae*
47 associated with recent widespread muskox mortalities in the Canadian Arctic. Can Vet J. 2015;56:560–
48 563.
49
50
51
52
53
54 18. Brooke CJ, Riley TV. *Erysipelothrix rhusiopathiae*: bacteriology, epidemiology and clinical
55 manifestations of an occupational pathogen. J Med Microbiol. 1999;48:789–799.
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 19. Ogawa Y, Ooka T, Shi F, Ogura Y, Nakayama K, Hayashi T, et al. The genome of *Erysipelothrix*
5
6 *rhusiopathiae*, the causative agent of swine erysipelas, reveals new insights into the evolution of
7
8 firmicutes and the organism's intracellular adaptations. J Bacteriol. 2011;193:2959–2971.
9
10
11 20. Verbarg S, Göker M, Scheuner C, Schumann P, Stackebrandt E. The Families Erysipelotrichaceae
12
13 emend., Coprobacillaceae fam. nov., and Turicibacteraceae fam. nov. In: Rosenberg E, DeLong EF, Lory S,
14
15 Stackebrandt E, Thompson F, editors. The Prokaryotes. Berlin: Springer; 2014. p. 79–105.
16
17
18 21. Takahashi T, Fujisawa T, Tamura Y, Suzuki S, Muramatsu M, Sawada T, et al. DNA relatedness among
19
20 *Erysipelothrix rhusiopathiae* strains representing all twenty-three serovars and *Erysipelothrix*
21
22 *tonsillarum*. Int J Syst Bacteriol. 1992;42:469–473.
23
24
25 22. Verbarg S, Rheims H, Emus S, Frühling A, Kroppenstedt RM, Stackebrandt E, et al. *Erysipelothrix*
26
27 *inopinata* sp. nov., isolated in the course of sterile filtration of vegetable peptone broth, and description
28
29 of Erysipelotrichaceae fam. nov. Int J Syst Evol Microbiol. 2004;54:221–225.
30
31
32 23. Bang B-H, Rhee M-S, Chang D-H, Park D-S, Kim B-C. *Erysipelothrix larvae* sp. nov., isolated from the
33
34 larval gut of the rhinoceros beetle, *Trypoxylus dichotomus* (Coleoptera: Scarabaeidae). Antonie Van
35
36 Leeuwenhoek. 2015;107:443–451.
37
38
39 24. Takahashi T, Sawada T, Muramatsu M, Tamura Y, Fujisawa T, Benno Y, et al. Serotype, antimicrobial
40
41 susceptibility, and pathogenicity of *Erysipelothrix rhusiopathiae* isolates from tonsils of apparently
42
43 healthy slaughter pigs. J Clin Microbiol. 1987;25:536–539.
44
45
46 25. Takahashi T, Fujisawa T, Umeno A, Kozasa T, Yamamoto K, Sawada T. A taxonomic study on
47
48 *Erysipelothrix* by DNA-DNA hybridization experiments with numerous strains isolated from extensive
49
50 origins. Microbiol Immunol. 2008;52:469–478.
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 26. Kucsera G. Proposal for standardization of the designations used for serotypes of *Erysipelothrix*
5
6
7 *rhusiopathiae* (Migula) Buchanan. Int J Syst Bacteriol. 1973;23:184–188.
8
9
10 27. Pal N, Bender JS, Opriessnig T. Rapid detection and differentiation of *Erysipelothrix* spp. by a novel
11
12 multiplex real-time PCR assay. J Appl Microbiol. 2010;108:1083–1093.
13
14
15 28. King SJ, Leigh JA, Heath PJ, Luque I, Tarradas C, Dowson CG, et al. Development of a multilocus
16
17 sequence typing scheme for the pig pathogen *Streptococcus suis*: identification of virulent clones and
18
19 potential capsular serotype exchange. J Clin Microbiol. 2002;40:3671–3680.
20
21
22 29. Ingebritson AL, Roth JA, Hauer PJ. *Erysipelothrix rhusiopathiae*: association of Spa-type with serotype
23
24 and role in protective immunity. Vaccine. 2010;28:2490–2496.
25
26
27 30. Wang Q, Chang BJ, Riley TV. *Erysipelothrix rhusiopathiae*. Vet Microbiol. 2010;140:405–417.
28
29
30 31. Mell JC, Redfield RJ. Natural competence and the evolution of DNA uptake specificity. J Bacteriol.
31
32 2014;196:1471–1483.
33
34
35 32. Laurenceau R, Péhau-Arnaudet G, Baconnais S, Gault J, Malosse C, Dujeancourt A, et al. A type IV
36
37 pilus mediates DNA binding during natural transformation in *Streptococcus pneumoniae*. PLoS Pathog.
38
39 2013;9:e1003473.
40
41
42 33. Edwards DJ, Holt KE. Beginner’s guide to comparative bacterial genome analysis using next-
43
44 generation sequence data. Microb Inform Exp. 2013;3:2.
45
46
47 34. Noguchi N, Sasatsu M, Takahashi T, Ohmae K, Terakado N, Kono M. Detection of plasmid DNA in
48
49 *Erysipelothrix rhusiopathiae* isolated from pigs with chronic swine erysipelas. J Vet Med Sci.
50
51 1993;55:349–350.
52
53
54 35. Eamens GJ, Forbes WA, Djordjevic SP. Characterisation of *Erysipelothrix rhusiopathiae* isolates from
55
56 pigs associated with vaccine breakdowns. Vet Microbiol. 2006;115:329–338.
57
58
59
60
61
62
63
64
65

- 1
2
3
4 36. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. ISME J.
5
6 2009;3:199–208.
7
8
9
10 37. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal
11
12 evolution in response to clinical interventions. Science. 2011;331:430–434.
13
14
15 38. Harris SR, Clarke IN, Seth-Smith HMB, Solomon AW, Cutcliffe LT, Marsh P, et al. Whole-genome
16
17 analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by
18
19 current clinical typing. Nat Genet. 2012;44:413–419, S1.
20
21
22
23 39. Grad YH, Kirkcaldy RD, Trees D, Dordel J, Harris SR, Goldstein E, et al. Genomic epidemiology of
24
25 *Neisseria gonorrhoeae* with reduced susceptibility to cefixime in the USA: a retrospective observational
26
27 study. Lancet Infect Dis. 2014;14:220–226.
28
29
30
31 40. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis
32
33 of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucl Acids Res.
34
35 2015;43:e15.
36
37
38
39 41. Darch SE, McNally A, Harrison F, Corander J, Barr HL, Paszkiewicz K, et al. Recombination is a key
40
41 driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic
42
43 fibrosis infection. Sci Rep. 2015;5:7649.
44
45
46
47 42. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, et al. Patterns of genome
48
49 evolution that have accompanied host adaptation in *Salmonella*. Proc Natl Acad Sci USA. 2015;112:863–
50
51 868.
52
53
54
55 43. den Bakker HC, Didelot X, Fortes ED, Nightingale KK, Wiedmann M. Lineage specific recombination
56
57 rates and microevolution in *Listeria monocytogenes*. BMC Evolutionary Biology. 2008;8:277.
58
59
60
61
62
63
64
65

- 1
2
3
4 44. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method for improved
5
6 phylogenetic and taxonomic placement of microbes. Nat Commun. 2013;4:2304.
7
8
9
10 45. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype
11
12 data. PLoS Genet. 2012;8:e1002453.
13
14
15 46. Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic
16
17 networks. Mol Biol Evol. 2004;21:255–265.
18
19
20 47. Harada T, Ogawa Y, Eguchi M, Shi F, Sato M, Uchida K, et al. Phosphorylcholine and SpaA, a choline-
21
22 binding protein, are involved in the adherence of *Erysipelothrix rhusiopathiae* to porcine endothelial
23
24 cells, but this adherence is not mediated by the PAF receptor. Vet Microbiol. 2014;172:216–222.
25
26
27 48. Borrathybay E, Gong F-J, Zhang L, Nazierbieke W. Role of surface protective antigen A in the
28
29 pathogenesis of *Erysipelothrix rhusiopathiae* strain C43065. J Microbiol Biotechnol. 2015;25:206–216.
30
31
32 49. To H, Nagai S. Genetic and antigenic diversity of the surface protective antigen proteins of
33
34 *Erysipelothrix rhusiopathiae*. Clin Vaccine Immunol. 2007;14:813–820.
35
36
37 50. Chooromoney KN, Hampson DJ, Eamens GJ, Turner MJ. Analysis of *Erysipelothrix rhusiopathiae* and
38
39 *Erysipelothrix tonsillarum* by multilocus enzyme electrophoresis. J Clin Microbiol. 1994;32:371–376.
40
41
42 51. Eriksson H, Jansson DS, Johansson K-E, Båverud V, Chirico J, Aspán A. Characterization of
43
44 *Erysipelothrix rhusiopathiae* isolates from poultry, pigs, emus, the poultry red mite and other animals.
45
46 Vet Microbiol. 2009;137:98–104.
47
48
49 52. Pan X, Yang Y, Zhang J-R. Molecular basis of host specificity in human pathogenic bacteria. Emerg
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

53. Hotchkiss EJ, Hodgson JC, Lainson FA, Zadoks RN. Multilocus sequence typing of a global collection of *Pasteurella multocida* isolates from cattle and other host species demonstrates niche association. BMC Microbiol. 2011;11:115.
54. Delannoy CMJ, Crumlish M, Fontaine MC, Pollock J, Foster G, Dagleish MP, et al. Human *Streptococcus agalactiae* strains in aquatic mammals and fish. BMC Microbiol. 2013;13:41.
55. Kirzinger MWB, Stavrinides J. Host specificity determinants as a genetic continuum. Trends Microbiol. 2012;20:88–93.
56. Opriessnig T, Shen HG, Bender JS, Boehm JR, Halbur PG. *Erysipelothrix rhusiopathiae* isolates recovered from fish, a harbour seal (*Phoca vitulina*) and the marine environment are capable of inducing characteristic cutaneous lesions in pigs. J Comp Pathol. 2013;148:365–372.
57. Wellmann GG. Summaries of experiments in swine erysipelas in Germany. JAVMA. 1955;127:331–333.
58. Muellner P, Stärk KDC, Dufour S, Zadoks RN. “Next-generation” surveillance: an epidemiologists’ perspective on the use of molecular information in food safety and animal health decision-making. Zoonoses Public Health. 2015;e-pub ahead of print.
59. Brazelton WJ, Morrill PL, Szponar N, Schrenk MO. Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs. Appl Environ Microbiol. 2013;79:3906–3916.
60. Wood RL. A selective liquid medium utilizing antibiotics for isolation of *Erysipelothrix insidiosa*. Am J Vet Res. 1965;26:1303–1308.
61. Bender JS, Kinyon JM, Kariyawasam S, Halbur PG, Opriessnig T. Comparison of conventional direct and enrichment culture methods for *Erysipelothrix* spp. from experimentally and naturally infected swine. J Vet Diagn Invest. 2009;21:863–868.

- 1
2
3
4 62. Takeshi K, Makino S, Ikeda T, Takada N, Nakashiro A, Nakanishi K, et al. Direct and rapid detection by
5
6 PCR of *Erysipelothrix* sp. DNAs prepared from bacterial strains and animal tissues. J Clin Microbiol.
7
8 1999;37:4093–4098.
9
10
11 63. Smeds L, Künstner A. ConDeTri - a content dependent read trimmer for Illumina data. PLoS ONE.
12
13 2011;6:e26314.
14
15
16 64. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Oxford
17
18 University Press. 2013;arXiv:1303.3997.
19
20
21 65. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap:
22
23 evaluating next-generation sequencing alignment data. Bioinformatics. 2012;28:2678–2679.
24
25
26 66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
27
28 format and SAMtools. Bioinformatics. 2009;25:2078–2079.
29
30
31 67. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open
32
33 software for comparing large genomes. Genome Biol. 2004;5:R12.
34
35
36 68. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome
37
38 assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–477.
39
40
41 69. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies.
42
43 Bioinformatics. 2013;29:1072–1075.
44
45
46 70. Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering for error
47
48 correction in single-cell sequencing. BMC Genomics. 2013;14:1–11.
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 71. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. A post-assembly genome-
5
6 improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc.* 2012;7:1260–
7
8 1284.
9
10
11 72. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation
12
13 of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 2014;42:D206–214.
14
15
16 73. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucl Acids Res.*
17
18 2011;39:W347–52.
19
20
21 74. Kears M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an
22
23 integrated and extendable desktop software platform for the organization and analysis of sequence
24
25 data. *Bioinformatics.* 2012;28:1647–1649.
26
27
28 75. Sahl JW, Caporaso JG, Rasko DA, Keim P. The large-scale blast score ratio (LS-BSR) pipeline: a method
29
30 to rapidly compare genetic content between bacterial genomes. *PeerJ.* 2014;2:e332.
31
32
33 76. Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, et al. PanGP: A tool for quickly analyzing bacterial pan-
34
35 genome profile. *Bioinformatics.* 2014;30:1297–1299.
36
37
38 77. Petkau A, Stuart-Edwards M, Stothard P, Domselaar GV. Interactive microbial genome visualization
39
40 with GView. *Bioinformatics.* 2010;26:3125–3126.
41
42
43 78. Treangen TJ, Ondov BD, Koren S, Phillippy AM. Rapid core-genome alignment and visualization for
44
45 thousands of microbial genomes. *Genome Biol.* 2014;15:524.
46
47
48 79. Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinformatics.*
49
50 2014;47:11.12.1–11.12.34.
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 80. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, et al. Detection of
5
6 recombination events in bacterial genomes from large population samples. *Nucl Acids Res.* 2012;40:e6.
7
8
9
10 81. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees
11
12 made easy. *Nucleic Acids Res.* 2011;39:W475–478.
13
14
15 82. Lawson D, Hellenthal G, Falush D, Myers S. PaintMyChromosomes.com: fineSTRUCTURE v2 &
16
17 GLOBETROTTER. <http://www.paintmychromosomes.com> (2012). Accessed 14 April 2015.
18
19
20 83. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by
21
22 maximum likelihood. *Syst Biol.* 2003;52:696–704.
23
24
25
26 84. Rambaut A. Molecular evolution, phylogenetics and epidemiology: Path-O-Gen.
27
28 <http://tree.bio.ed.ac.uk/software/pathogen> (2007). Accessed 28 April 2015.
29
30
31 85. Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for
32
33 phylogenetic uncertainty. *Infect Genet Evol.* 2008;8:239–246.
34
35
36
37 86. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient
38
39 Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61:539–
40
41 542.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figures

Figure 1 *Erysipelothrix rhusiopathiae* pan-genome. Plots were generated in LS-BSR and visualized using PanGP. (A) Convergence of the core genome with concurrent accumulation of coding sequences in the pan-genome in relation to the number of genomes analyzed. (B) Number of unique coding sequences for each additional genome analyzed. Publicly available *E. rhusiopathiae* isolates and *de novo* assembled isolates whose average depth of coverage was greater than 15X were included in this analysis.

Figure 2 Relationship among the species of the genus *Erysipelothrix*. This phylogenetic tree is based on > 400 conserved bacterial protein sequences, generated using PhyloPhlAn. Members of other genera of the family Erysipelotrichaceae were used as the outgroup for rooting the tree. The scale bar represents the expected number of nucleotide substitutions per sequence position.

Figure 3 Population structure of *Erysipelothrix rhusiopathiae* based on phylogenetic inference. This tree is based on > 400 conserved bacterial protein sequences, generated using PhyloPhlAn. Other *Erysipelothrix spp.* were used as the outgroup for rooting the tree (clade collapsed). Concentric rings illustrate host species of origin (colored squares) and geographic origin (grey-scale). Arrows indicate isolates whose correct clade association is not resolved between the phylogenetic and chromosome painting approaches. Presence of the surface protective antigen type B (*SpaB*) gene was exclusive to Clade 1, while the *SpaA* gene was found among the two other clades and intermediate isolates.

Figure 4 Population structure of *Erysipelothrix rhusiopathiae* based on *in silico* chromosome painting. Tree and heat map illustrate the relatedness among *E. rhusiopathiae* isolates based on chromosome painting using ChromoPainter and model-based clustering using Bayesian Markov chain Monte Carlo (MCMC) analysis in fineSTRUCTURE. The color scale represents the number of 'chunks' shared between populations of donors (x-axis) and recipients (y-axis). Population subgroup assignment is shown in Figure

S4 in Additional file 5. Arrows indicate isolates whose correct clade association is not resolved between the phylogenetic and chromosome painting approaches.

Figure 5 Population structure of *Erysipelothrix rhusiopathiae* based on network analysis. A

phylogenetic network of *E. rhusiopathiae* was estimated using Neighbor-Net analysis as implemented in SplitsTree. Arrows indicate isolates whose correct clade association is not resolved between the phylogenetic and chromosome painting approaches.

Figure 6 Relationship among Clade 3 isolates. This maximum likelihood tree (mid-point rooted) was generated using PhyML based on the curated set of single nucleotide polymorphisms (SNPs) found to be outside recombinant segments as determined using Gubbins. Bootstrap values with >70% support are shown, in addition to support for clusters of isolates by PhyloPhlAn and fineSTRUCTURE.

Table 1 Recombination detected in the core genome of *Erysipelothrix rhusiopathiae*

	Clade 1	Clade 2	Clade 3
% of alignment implicated BNG	0	38	24
% of alignment implicated Gubbins	27	86	58
r/m	0.96	2.18	0.55

Percentage of the *Erysipelothrix rhusiopathiae* core genome found to have experienced recombination within each clade using BratNextGen (BNG) and Gubbins. Recombination to mutation (r/m) rates for each clade were estimated in Gubbins.

Figure 1

[Click here to download Figure Figure 1.tif](#)

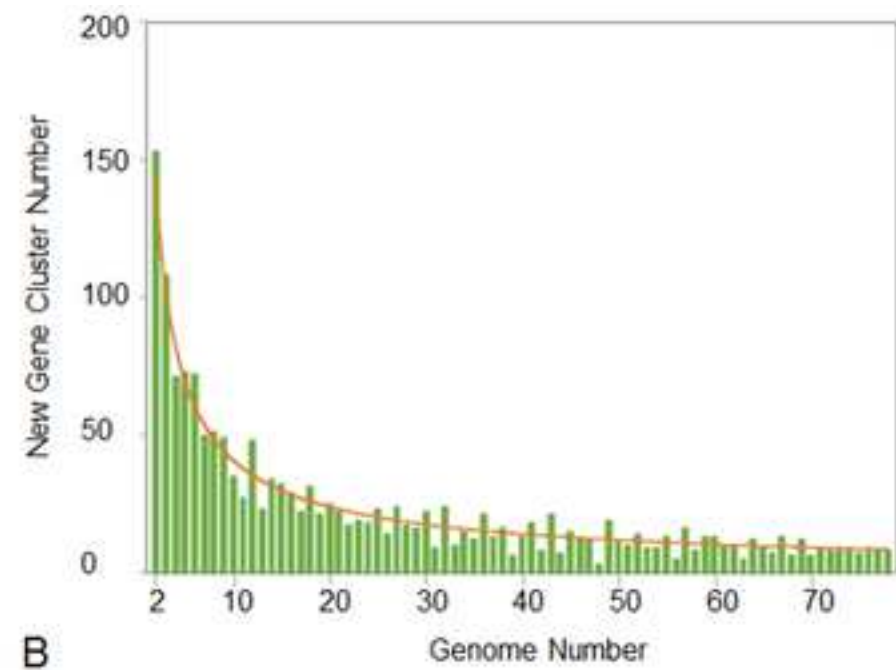
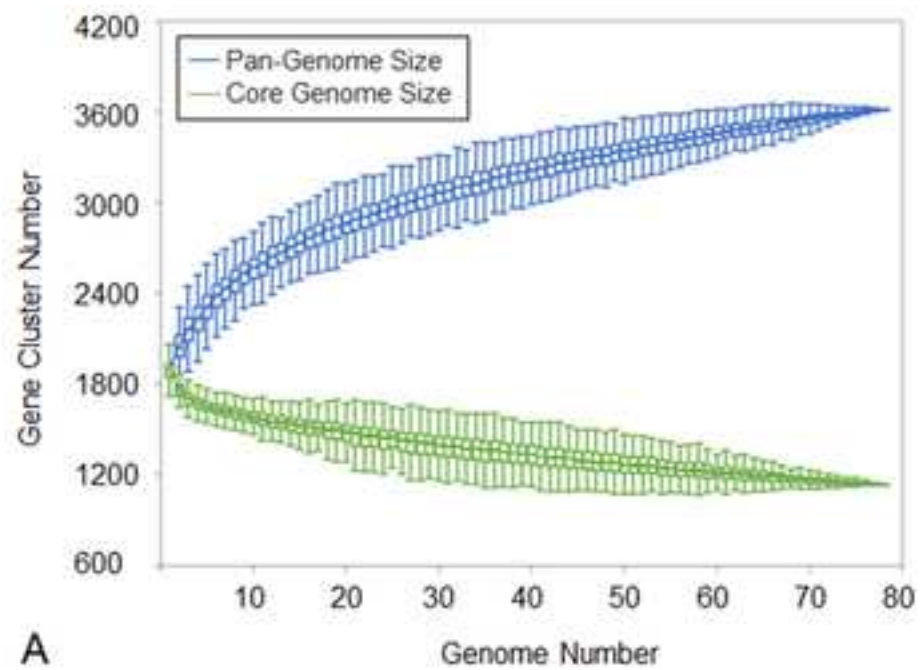


Figure 2

[Click here to download Figure Figure 2.tif](#)

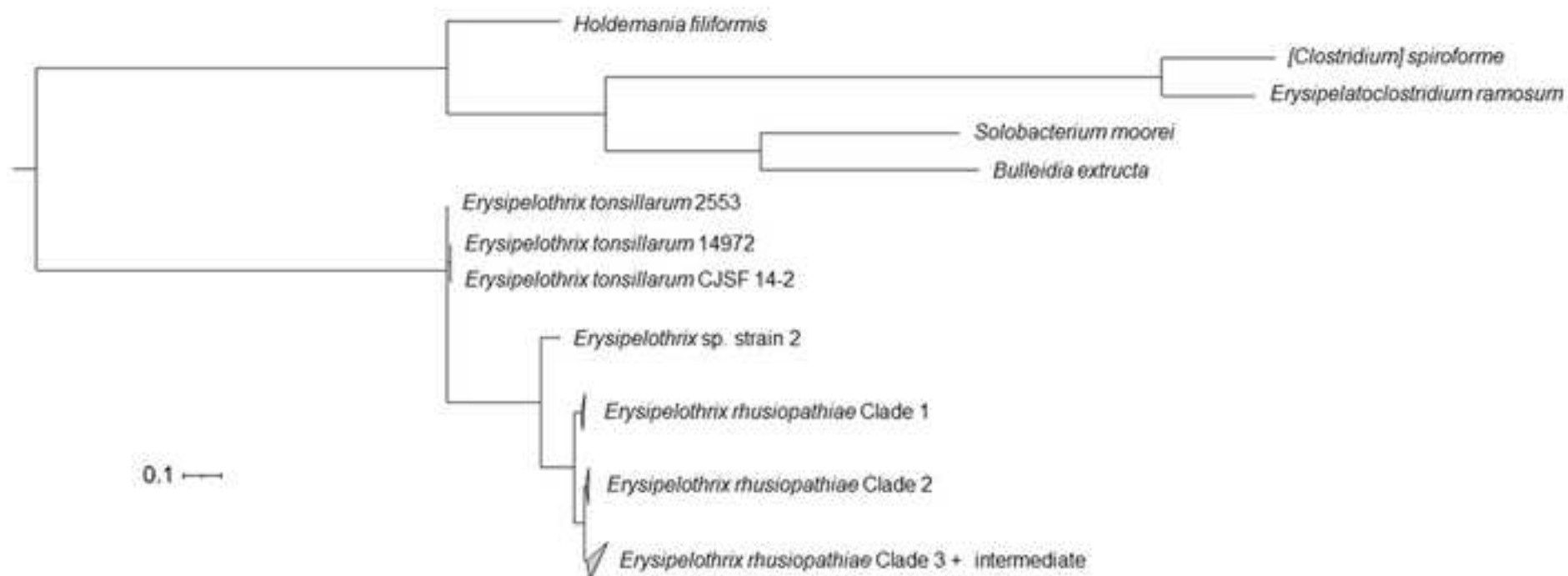


Figure 3

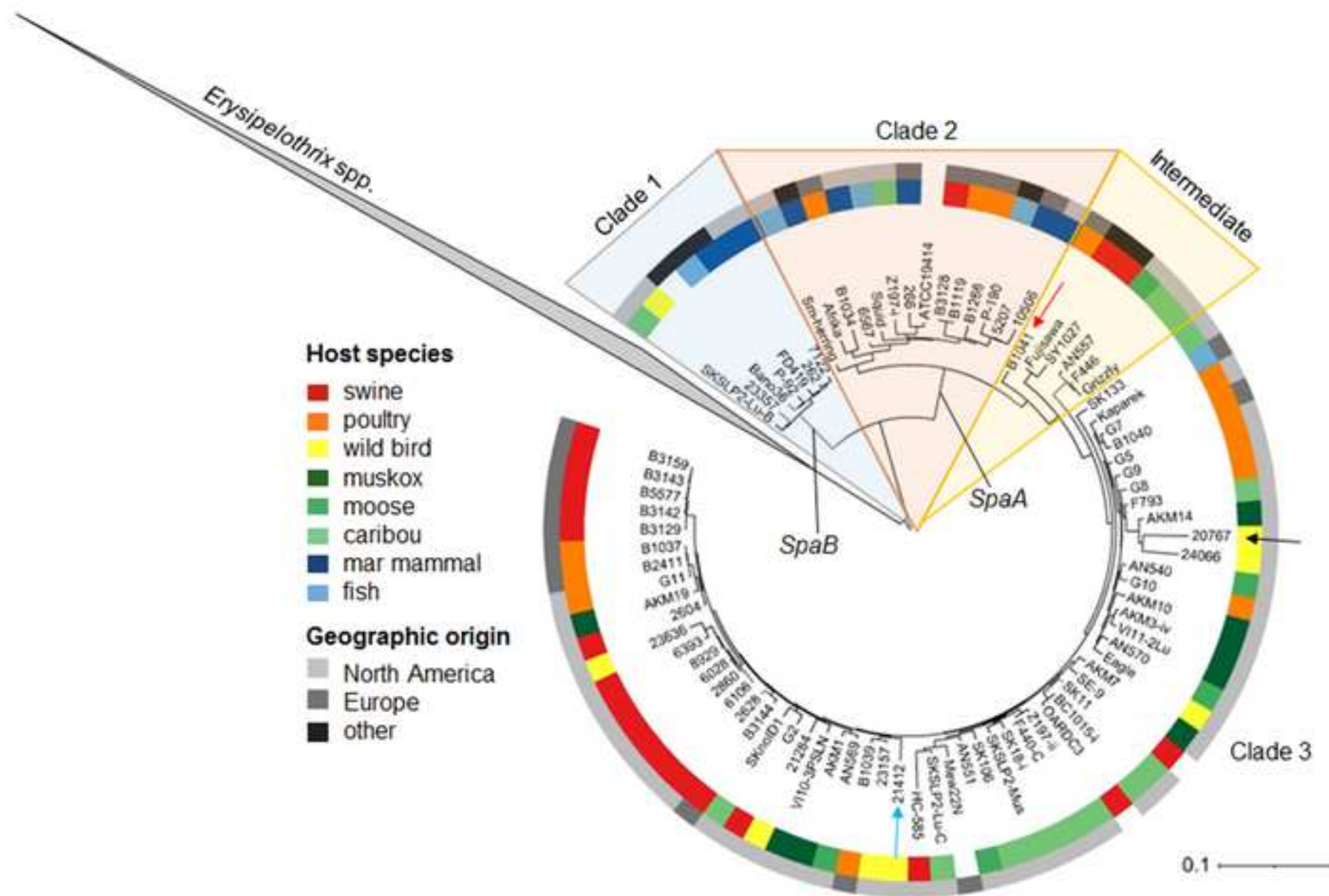


Figure 4

[Click here to download Figure Figure 4.tif](#)

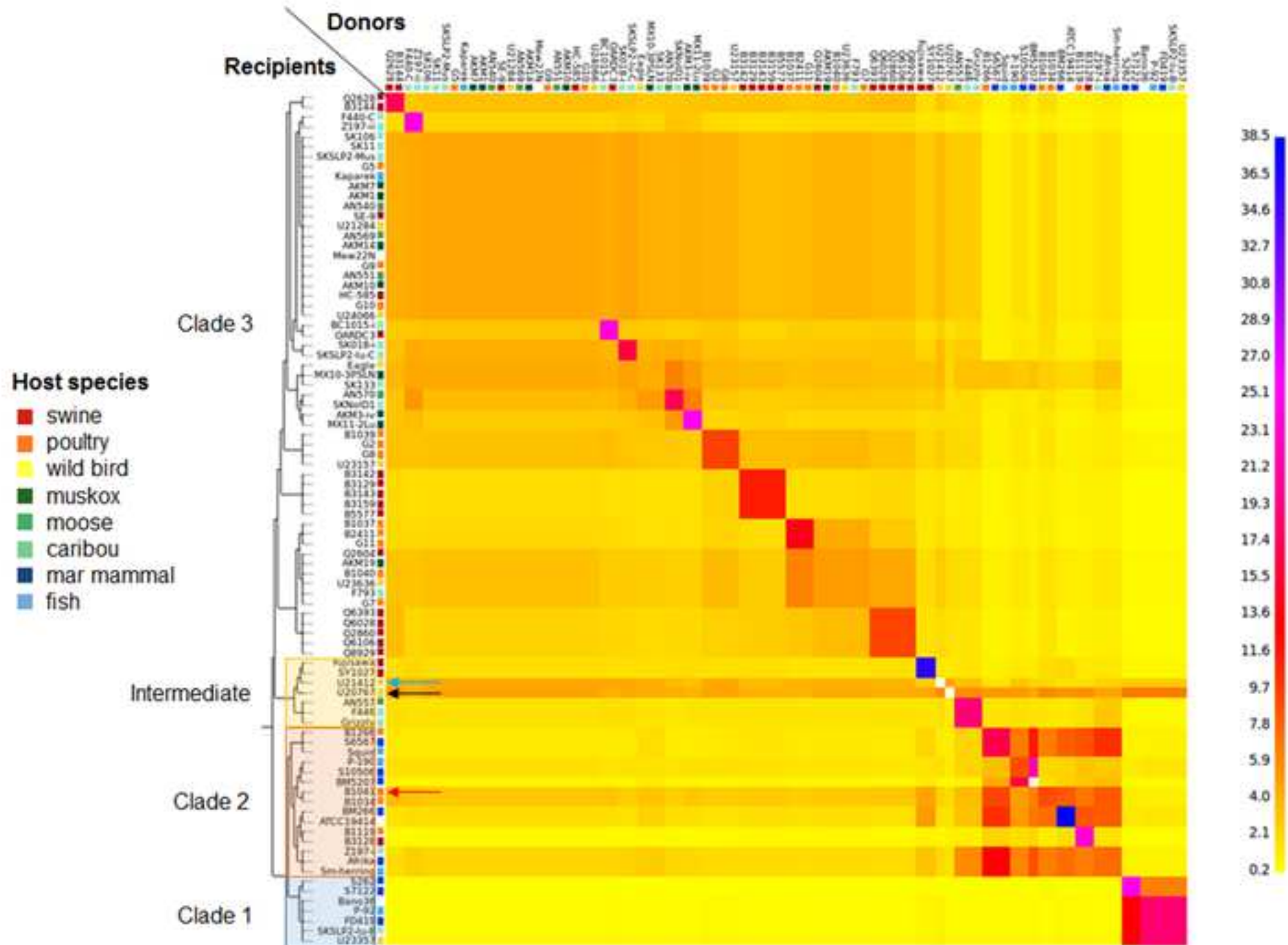
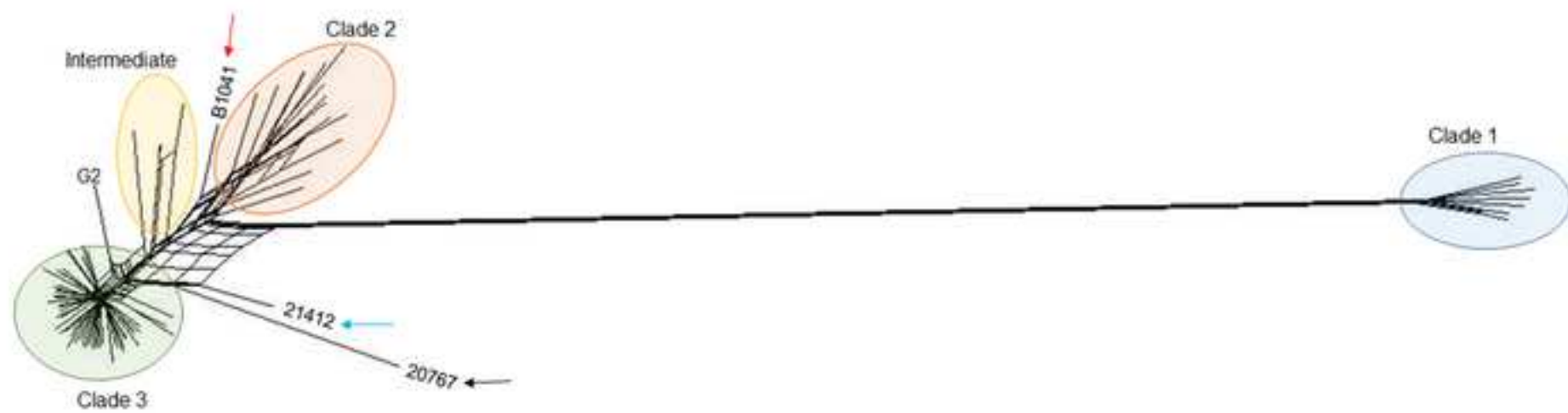
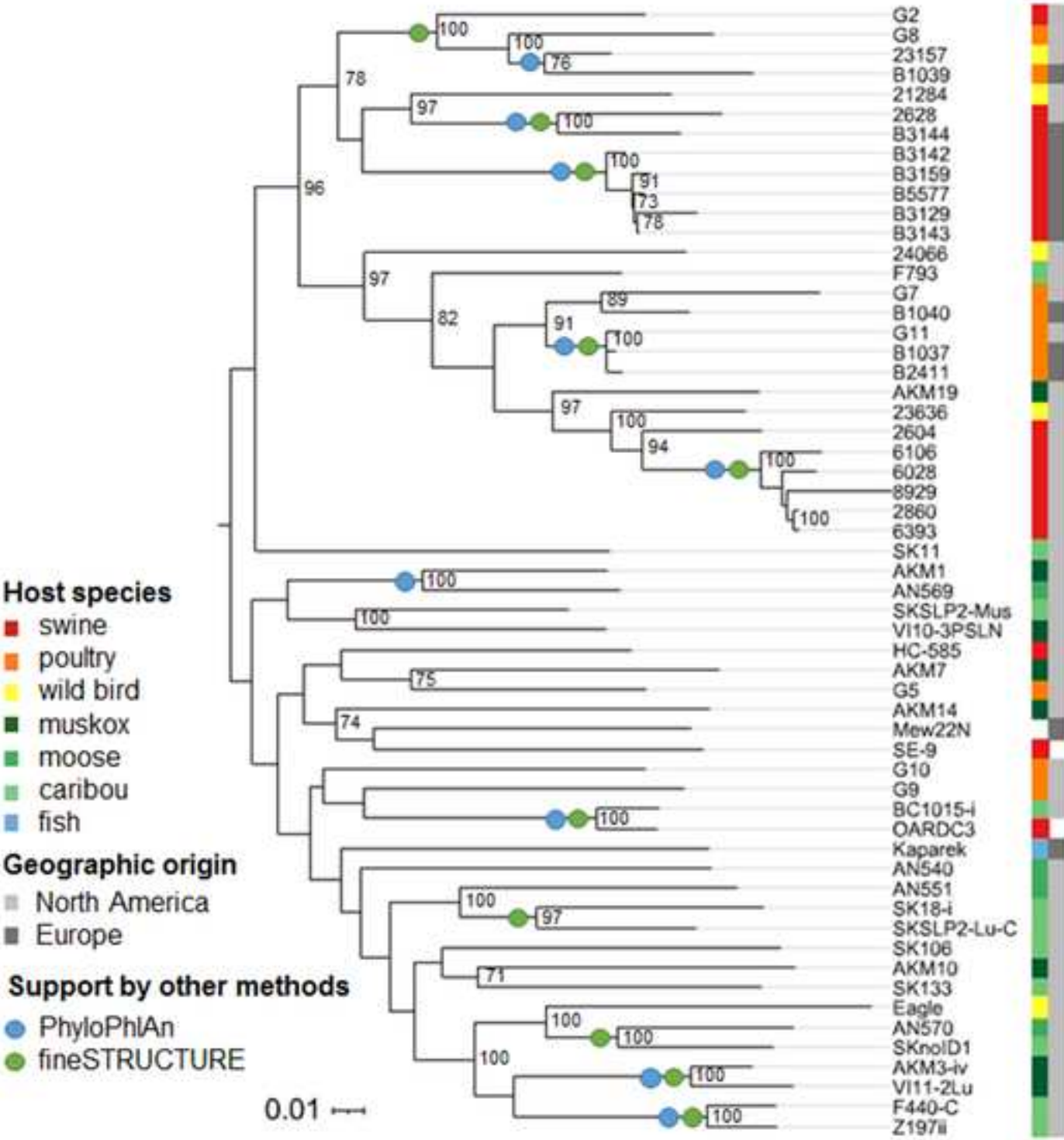
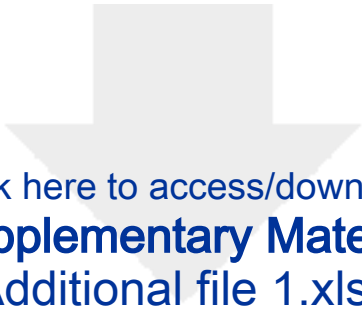


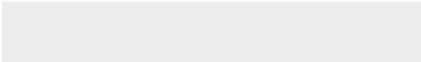

Figure 5






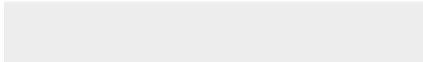



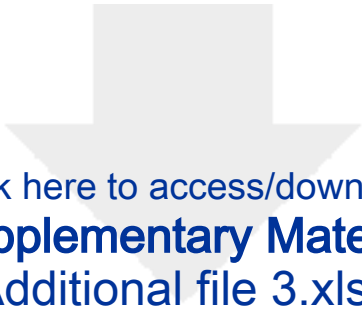
Click here to access/download
Supplementary Material
Additional file 1.xlsx



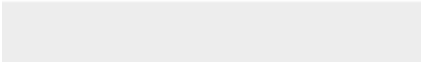




Click here to access/download
Supplementary Material
Additional file 2.pdf





Click here to access/download
Supplementary Material
Additional file 3.xlsx





Click here to access/download
Supplementary Material
Additional file 4.txt

